[18.02] Efficient GAN-Based Anomaly Detection

2019-01-26

工学系研究科 技術経営戦略学専攻 松尾研究室 修士2年
37-176839 田村 浩一郎

Paper Information

- [18.02] Efficient GAN-Based Anomaly Detection
  - Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, Vijay Ramaseshan Chandrasekhar
  - https://arxiv.org/abs/1802.06222
  - ICLR 2018 workshop
  - GANを用いた異常検知論文，AnoGANの改良版
    - GANを用いた教師なし異常検知
    - GANを学習する際にx->zのEncoderも同時に学習することによって，AnoGANにおいてコストの大きかった潜在空間におけるzの探索の手間を省いた

Outline

- Introduction & Related Work
- EfficientGAN
- Experiment
- Conclusion & Comment

Introduction & Related Work

- 異常検知は，製造業，医療，サイバーセキュリティなどにおいて重要な研究分野である
- GANによって，高次元でかつ複雑な通常データの分布のモデル化が成功している
- しかし，既存研究のGANを用いた異常検知では，潜在空間への写像が明示的ではなく，AnoGANなどの手法に見られるような潜在空間からのマッピングにおける探索が大きなコストだった
- そこで本研究は，以下を提案
  - GANの学習と同時にEncoderを学習することによって，潜在空間における探索を省略する
- 先行研究としては，
  - Deep autoencoder Gaussian mixture modelsなどが異常検知で用いられている
  - AnoGANはGANを用いた異常検知の先駆け

EfficientGAN

- GANを用いた異常検知モデルとして，AnoGANやADGANなどがあるが，推論時に潜在空間の探索を行うため，リアルタイムな処理では特に実用的ではなかった
  - GAN学習時において，画像を潜在空間にマッピングするEncoderを同時に学習し，推論時に利用することで既存手法の数百倍の速度を実現
  - 基本的なGANアーキテクチャは，BiGANを用いている
    - Adversarial Feature Learning
    - https://arxiv.org/abs/1605.09782

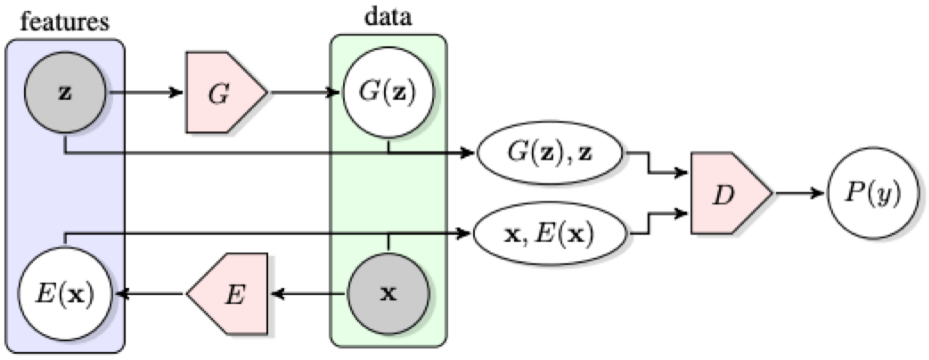

Figure 1: The structure of Bidirectional Generative Adversarial Networks (BiGAN).

EfficientGAN

- BiGANのように，GANにおける潜在空間への写像とは別に，Encoder E(x)を学習し，Discriminatorには画像と潜在変数(ノイズ)のペアを入力とすることで学習する
  - (x, z)と画像とノイズのペアで損失関数を定義する

$$V(D,E,G) = \mathbb{E}_{x\sim p_X}\left[\mathbb{E}_{z\sim p_E(\cdot|x)}\left[\log D(x,z)\right]\right] + \mathbb{E}_{z\sim p_Z}\left[\mathbb{E}_{x\sim p_G(\cdot|z)}\left[1-\log D(x,z)\right]\right].$$

    - (x, E(x)): 実際の画像と，encodeされたノイズのペア
    - (G(z), z): Generatorによって生成された画像と，一様乱数に従うノイズのペア

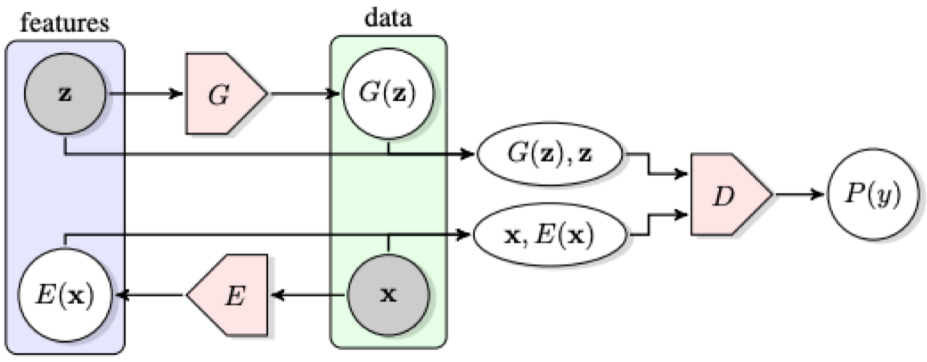

Figure 1: The structure of Bidirectional Generative Adversarial Networks (BiGAN).

EfficientGAN

- 異常検知における判定は，学習したEncoderがあるので，探索する必要がない
  - E(x)がAnoGANにおける探索された潜在変数z_γとして良い
  - 一回の推論によって，AnoGANと同じようにAnomaly scoreを用いて判定することができる

$$
\begin{aligned}
A(x) &= \alpha L_G(x) + (1-\alpha) L_D(x) \\
L_G(x) &= \|x - G(E(x))\|_1 \\
L_D(x) &= \|f_D(x, E(x)) - f_D(G(E(x)), E(x))\|_1
\end{aligned}
$$

Experiment

- MNIST, KDDCUP99で実験を行なった
- MNIST
  - MNISTにおける異なる10の数値の1つを異常とし，他を正常とした
  - 正常データの8割をtraining, 正常データ2割と異常データをtestとした
  - VAEとAnoGANで比較
  - なお，Anomaly Scoreを計算する際に，discriminatorの中間層からの出力を用いる場合も実験(そっちの方が好パフォーマンス)

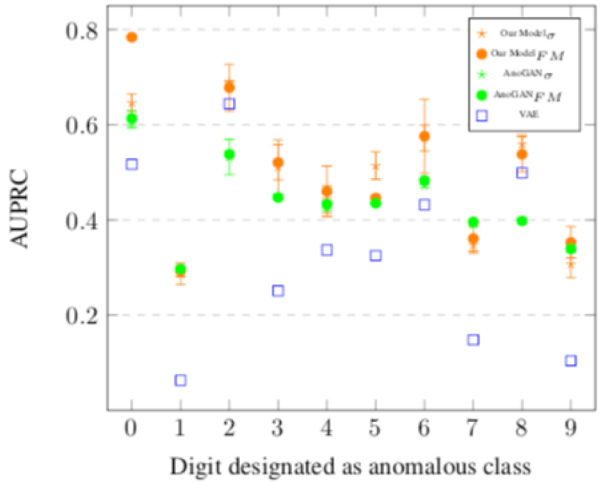

Figure 1: Performance on MNIST measured by the area under the precision-recall curve. Best viewed in color. VAE data were obtained from (An & Cho, 2015). Error bars show variation across 3 random seeds. FM and σ respectively denote feature-matching and cross-entropy variants of L_D used in the anomaly score.

Experiment

- KDDCUP99
  - KDDCUP99の10%datasetの値のうち，上位20%を異常とした
  - 正常データの50%をtraining，正常データ50%と異常データをtestとする
  - OC-SVM, DSEBM，DAGMM, AnoGANと比較
  - なお，Anomaly Scoreを計算する際に，discriminatorの中間層からの出力を用いる場合も実験(そっちの方が好パフォーマンス)
  - AnoGANの推論速度を700~900倍上回る

Table 1: Performance on the KDD99 dataset. Values for OC-SVM, DSEBM, DAGMM values were obtained from (Zhai et al., 2016; Bo Zong, 2018). Values for AnoGAN and our model are derived from 10 runs.

Model	Precision	Recall	F1
OC-SVM	0.7457	0.8523	0.7954
DSEBM-r	0.8521	0.6472	0.7328
DSEBM-e	0.8619	0.6446	0.7399
DAGMM-NVI	0.9290	0.9447	0.9368
DAGMM	**0.9297**	0.9442	0.9369
AnoGAN$_{FM}$	0.8786 ± 0.0340	0.8297 ± 0.0345	0.8865 ± 0.0343
AnoGAN$_{\sigma}$	0.7790 ± 0.1247	0.7914 ± 0.1194	0.7852 ± 0.1181
Our Model$_{FM}$	0.8698 ± 0.1133	0.9523 ± 0.0224	0.9058 ± 0.0688
Our Model$_{\sigma}$	0.9200 ± 0.0740	$\mathbf{0.9582 \pm 0.0104}$	$\mathbf{0.9372 \pm 0.0440}$

Conclusion & Comment

どんなもの？

推論速度の速いGAN-baseでの異常検知

どうやって有効だと証明した？

MNISTとKDDCUP99に対してSOTA．画像データ，配列データ共に良い性能．

先行研究との差分は？

AnoGANの改良版．Encoderを学習することによって，推論速度を大幅にはやめた

議論点

- パラメタが多い分チューニングが難しそう

技術や手法のキモは？

Encode->decode型のGANを用いて，reconst errorとdiscriminator 出力を組み合わせた指標で異常度を測る．Encoderの関数を学習することで，潜在空間におけるzの探索をする必要がなくなった

次に読むべき論文は？

1. Anomaly Detection with Generative Adversarial Networks

2. AVID: Adversarial Visual Irregularity Detection